

PACKET SWITCH REALIZING TRANSMISSION WITH NO PACKET DELAY

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates to a packet switch for use in a packet switching system communication network and, more particularly, to a large-scale packet switch having unit switches connected in multi-stages.

DESCRIPTION OF THE RELATED ART

In a high-speed packet switching system communication network, a large-scale packet switch handling more than several hundreds of lines is realized by connecting a medium- or small-scale packet switches in multi-stages. As an example of a packet switch having such a multi-stage connection structure, a three-stage connection packet switch is shown in Fig. 13.

The packet switch illustrated in Fig. 13 is composed of connected unit switches 1311 and 1312 at the first stage, unit switches 1321 and 1322 at the second stage and unit switches 1331 and 1332 at the third stage. Here, as communication paths, there exist a plurality of paths made by a combination between arbitrary input port and output port.

For accommodating connection-oriented traffic as in ATM (Asynchronous Transfer Mode), selection of a path

to be used is made at the setting of a call. On this occasion, for preventing internal blocking (a state where although a free band exists at a final output port, path setting is impossible within the unit switches 1321 and 1322 at the second stage), it is necessary to set paths so as not to cause a deviation in load.

For accommodating connectionless IP traffic, a packet belonging to the same flow should pass through the same path. There therefore arises the need of path selection every time flow starts similarly to the case of connection-oriented traffic.

Then, for either traffic, path selection involving no internal blocking needs centralized control by the entire switch, which results in sacrificing extensibility.

On the other hand, there is a method of preventing internal blocking by, at the outputting of packets from unit switches at the first stage to unit switches at the second stage, evenly distributing the packets to the unit switches at the second stage to macroscopically uniform a load of a unit switch at each second stage. At a multi-stage connection packet switch shown in Fig. 14, switching is conducted according to a destination of a packet at unit switches 1421 and 1422 at the second stage and at a unit switch 1431 at the final stage after returning.

According to such a switching method, however,

00540990 033100

5

15

20

25

the unit switches 1321 and 1322 at the second stage irrespective of their destinations. At the unit switches 1321 and 1322 at the second stage and the unit switches 1331 and 1332 at the third stage, the packets are sent to output ports corresponding to their destinations.

At the output ports of the unit switches 1331 and 1332 at the third stage, the packets are temporarily held for delay. Then, the packets are sequentially output to the output ports, starting with a packet which has passed a time period longer than a predetermined time after the application to the unit switches 1311 and 1312 at the first stage (a maximum delay time required for passing from the unit switches 1311 and 1312 at the first stage to the unit switches 1331 and 1332 at the third stage). The foregoing processing enables packet sequencing.

Another conventional packet switch using a time stamp is, for example, a packet switch recited in Japanese Patent Laying-Open (Kokai) No. heisei 6-6370. According to the technique disclosed in the literature, a unit switch at each first stage first distributes output cells to unit switches at the second stage according to output conditions of cells. Then, a unit switch at the third stage selects, in one clock cycle and for each output line, a cell whose immediately preceding cell is already output from each logic buffer provided corresponding to each unit switch at the first

5 Conventional packet switches using a time stamp,
however, output packets with a delay of a predetermined
time in order to sequence packets arriving at unit
switches at the final stage as described above. As a
result, the switches have a shortcoming that a time
10 required for packet transmission can not be made shorter
than a maximum delay time required for a packet to pass
from a unit switch at the first stage to a unit switch
at the third stage.

20 SUMMARY OF THE INVENTION

According to one aspect of the invention, a packet switch formed by connecting unit switches in

multi-stages, wherein

a unit switch at the first stage assigns a sequence number to an input packet according to a destination of the packet and distributes and sends out the packet to a unit switch at a succeeding stage, and

a unit switch at the final stage sequences and outputs a packet received from a unit switch at a preceding stage according to a sequence number assigned to the packet.

In the preferred construction, the unit switch at the first stage assigns, to an input packet, a sequence number set for each combination of a unit switch which has received input of the packet and a unit switch which finally outputs the packet, as well as assigning identification information about its own switch which is a unit switch having received input of the packet.

In another preferred construction, the total number of sequence numbers to be assigned by the unit switch at the first stage to an input packet is set based on a maximum value of a queuing delay at a unit switch at the succeeding stage and the number of input and output ports of the unit switch in question at the succeeding stage.

In another preferred construction, the unit switch at the first stage assigns, to an input packet, a sequence number set for each combination of a unit switch which has received input of the packet and a unit

00700"06604560

switch which finally outputs the packet, as well as assigning identification information about its own switch which is a unit switch having received input of the packet, and

5 the unit switch at the final stage includes queues provided for the respective unit switches at the first stage which receive input of packets and slotted on a packet basis, based on the identification information and the sequence number assigned to a packet
10 arriving from a unit switch at the preceding stage, writes the packet in question into a corresponding slot of corresponding one of the queues, and sequentially reads and outputs the packets written in the queues according to the order of the sequence numbers.

15 In another preferred construction, the unit switch at the final stage determines an initial value of a read pointer for the queue based on a sequence number of a packet received first.

20 In another preferred construction, when a packet is read from a slot indicated by a read pointer in the queue or when no packet arrives at the slot indicated by the read pointer in the queue for a predetermined time period, the unit switch at the final stage updates the read pointer in question.

25 In another preferred construction, the unit switch at the final stage determines an initial value of a read pointer for the queue based on a sequence number

00540990-033100

of a packet received first, and when a packet is read from a slot indicated by a read pointer in the queue or when no packet arrives at the slot indicated by the read pointer in the queue for a predetermined time period, updates the read pointer in question.

In another preferred construction, the unit switch comprises a $2N \times 2$ switch unit having a number $2N$ of input ports and a number N of output ports, a packet distribution unit for assigning, to an input packet, a sequence number according to a destination of the packet and distributing and sending out the packet to a unit switch at a succeeding stage, a packet alignment unit for sequencing, according to a sequence number assigned to a packet received from a unit switch at a preceding stage, a packet and transferring the sequenced packets to the $2N \times N$ switch unit, and a filter for distributing input packets into packets to be returned within its own switch and packets to be sent out to a unit switch at a succeeding stage and transferring the packets to be returned to the $2N \times N$ switch unit and transferring the packets to be sent to a switch at a succeeding stage to the packet distribution unit.

In another preferred construction, the packet distribution unit at the unit switch at the first stage assigns, to an input packet, a sequence number set for each combination of a unit switch which has received input of the packet and a unit switch which finally

00000"06604560

outputs the packet, as well as assigning identification information about its own switch which is a unit switch having received input of the packet.

In another preferred construction, the total number of sequence numbers to be assigned by the packet distribution unit in the unit switch at the first stage to an input packet is set based on a maximum value of a queuing delay at a unit switch at the succeeding stage and the number of input and output ports of the unit switch in question at the succeeding stage.

In another preferred construction, the unit switch at the final stage includes queues provided for the respective unit switches at the first stage which receive input of packets and slotted on a packet basis, based on the identification information and the sequence number assigned to a packet arriving from a unit switch at a preceding stage, the packet alignment unit writes the packet in question into a corresponding slot of corresponding one of the queues, and

the $2N \times N$ switch unit sequentially reads and outputs the packets written in the queues according to the order of the sequence numbers.

In another preferred construction, the unit switch at the final stage determines an initial value of a read pointer for the queue based on a sequence number of a packet received first.

In another preferred construction, when a packet

00750" 06604560

is read from a slot indicated by a read pointer in the queue or when no packet arrives at the slot indicated by the read pointer in the queue for a predetermined time period, the unit switch at the final stage updates the read pointer in question.

In another preferred construction, the unit switch at the final stage determines an initial value of a read pointer for the queue based on a sequence number of a packet received first, and when a packet is read from a slot indicated by a read pointer in the queue or when no packet arrives at the slot indicated by the read pointer in the queue for a predetermined time period, updates the read pointer in question.

Other objects, features and advantages of the present invention will become clear from the detailed description given herebelow.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood more fully from the detailed description given herebelow and from the accompanying drawings of the preferred embodiment of the invention, which, however, should not be taken to be limitative to the invention, but are for explanation and understanding only.

In the drawings:

Fig. 1 is a block diagram showing a structure of a packet switch according to one embodiment of the

present invention;

Fig. 2 is a block diagram showing a structure of a unit switch constituting the packet switch illustrated in Fig. 1;

5 Fig. 3 is a block diagram showing a structure of a $2N \times N$ switch unit in the present embodiment;

Fig. 4 is a diagram for use in explaining operation of a packet distribution unit in the present embodiment;

10 Fig. 5 is a diagram for use in explaining packet accumulation operation by a packet alignment unit in the present embodiment;

15 Fig. 6. is a diagram for use in explaining operation of outputting a packet from the packet alignment unit;

Fig. 7 is a diagram showing how a packet is abandoned due to a buffer overflow of a unit switch while the packet passes through a packet switch;

20 Fig. 8 is a flow chart showing operation of determining an initial value of a read pointer for use in reading a packet from a queue by the packet alignment unit in the present embodiment;

25 Fig. 9 is a flow chart showing operation of updating an initial value of a read pointer for use in reading a packet from a queue by the packet alignment unit in the present embodiment;

Fig. 10 is a schematic diagram for use in

00750" 06604560

explaining a manner of accommodating a line whose speed is higher than that of an input port of a unit switch in the packet switch of the present embodiment;

Fig. 11 is a diagram showing an example of a structure of a packet switch in which $N^2 \times N^2$ packet switches as a unit switch are connected in multi-stages to extend to $N^3 \times N^3$;

Fig. 12 is a block diagram showing a structure of a unit switch constituting the packet switch illustrated in Fig. 11;

Fig. 13 is a block diagram showing a structure of a three-stage connection packet switch;

Fig. 14 is a diagram showing how a packet passes through a multi-stage connection packet switch.

DESCRIPTION OF THE PREFERRED EMBODIMENT

The preferred embodiment of the present invention will be discussed hereinafter in detail with reference to the accompanying drawings. In the following description, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be obvious, however, to those skilled in the art that the present invention may be practiced without these specific details. In other instance, well-known structures are not shown in detail in order to unnecessary obscure the present invention.

Fig. 1 is a block diagram showing a structure of

a packet switch according to one embodiment of the present invention. With reference to Fig. 1, the packet switch of the present embodiment is constituted by two-stage connection of a number N of unit switches 101 each equivalent to $2N \times 2N$. A switch having such a connection form as illustrated in the figure is referred to as a folded-type multi-stage switch.

In the present embodiment, description will be made of a case where the present invention is applied to such a folded-type multi-stage switch as illustrated in the figure.

In a folded-type multi-stage switch, connecting unit switches each equivalent to $2N \times 2N$ in two stages forms a packet switch equivalent to $N^2 \times N^2$. Also, connecting the same in three stages forms a packet switch equivalent to $N^3 \times N^3$. The number of unit switches required for a packet switch of such a structure is $2N$ when the packet switch equivalent to $N^2 \times N^2$ is structured and $3N^2$ when the packet switch equivalent to $N^3 \times N^3$ is formed.

On the other hand, in a case where a packet switch is formed by square lattice expansion, the number of unit switches required for a packet switch equivalent to $N^2 \times N^2$ is N^2 , while the number of unit switches required for a packet equivalent to $N^3 \times N^3$ is N^4 . A folded-type multi-stage switch is accordingly allowed to reduce the number of necessary unit switches to be smaller than

that required for a packet switch formed by square lattice expansion.

In the folded-type multi-stage switch of the present embodiment, a series of sequence numbers (SN) is assigned, at each input port, to packets belonging to the same combination of an input side unit switch and an output side unit switch. Then, at a unit switch at the final stage, the packets are sequenced according to the SN numbers. The sequenced packets are sent to output ports corresponding to their destinations.

In the following, the present embodiment will be described assuming that a packet has a fixed length. When a packet of a variable length is handled, it is only necessary to add, to an input part of a packet at a unit switch at the first stage, a structure for fragmenting a packet of a variable length into packets of a fixed length and add, to an output part of the packet at a unit switch at the final stage, a structure for bringing packets of a fixed length together into a packet of a variable length. It should be clearly understood that adding these structures for fragmenting a packet of a variable length into packets of a fixed length and bringing packets of a fixed length together into a packet of a variable length is not a technical characteristic of the present embodiment.

Structure of the unit switch 101 equivalent to $2N \times 2N$ according to the present embodiment is shown in

09540990"033100
NOTED"06604560

Fig. 2. With reference to Fig. 2, the unit switch includes a filter unit 210, a $2N \times N$ switch unit 220, a packet distribution unit 230 and a packet alignment unit 240. Of these units, the packet distribution unit 230 and the packet alignment unit 240 are circuits for extension, that is, for connecting unit switches in multi-stages.

In the above described structure, the filter unit 210 allots an input packet according to its destination. More specifically, when an input packet is to be returned in its own switch, the unit 210 transfers the input packet in question to the $2N \times N$ switch unit 220 and when the packet is to be sent to other unit switch, transfers the same to the packet distribution unit 230.

The $2N \times N$ switch unit 220 has a number $2N$ of input ports and a number N of output ports. The switch unit 220 has input ports double the number of output ports because when a multi-stage connection is made, it is necessary to process both a packet which is to return in its own switch and a packet received from other unit switch. When the unit switch operates singly, the switch unit 220 operates as an $N \times N$ switch because no packet is applied from other unit switch.

Structure of the $2N \times N$ switch unit 220 is shown in Fig. 3.

With reference to Fig. 3, the $2N \times N$ switch unit 220 is an output buffer switch. In the illustrated

00750" 06604560

5

10

15

20

25

own switch. On this occasion, a series of sequence numbers (SN) is assigned to packets having the same combination of a unit switch which receives input of the packet (unit switch at the first stage) and a unit switch which outputs the packet (unit switch at the final stage). In practice, since the unit switch at the first stage is its own switch, packets with the same series of sequence numbers (SN) will arrive at each unit switch at the final stage.

In addition, assigned to all the packets is identification information indicating through which unit switch 101 at the first stage a packet in question is applied (e.g. a number of a unit switch of its own). These arrangement enables referring, at a unit switch at the final stage, to a sequence number (SN) and identification information about a unit switch at the first stage assigned to an arriving packet to lead to alignment of packets for each unit switch at the first stage which has received input of the packet in question and according to the order of application to the unit switch in question at the first stage.

As a packet distribution method, there is such a system of shifting a connection relation between an input and an output one by one irrespectively of a destination of a packet as barrel shifter, for example. An example of distribution and output in this case is shown in Fig. 4. With reference to Fig. 4, packets

00540990-033400

applied to the packet distribution unit 230, those bound for a unit switch #1 (packets with slanting lines) and those bound for a unit switch #N (packets with no slanting lines), are separately assigned sequence numbers (SN) 1-3 and are output with a connection relationship between an input and an output shifted one by one.

Although the above-described distribution method is considered to achieve sufficient performance, the performance can be further improved by distributing packets using a sequence number (SN) sent from a unit switch at a succeeding stage to a unit switch at a preceding switch and feedback information regarding the amount of queuing. It is, for example, possible to adopt a method of preferentially sending a packet to a unit switch whose queuing amount is smaller.

The packet alignment unit 240 aligns packets received from a unit switch at the preceding stage and transfers the aligned packets to the $2N \times N$ switch unit 220. Structure of the packet alignment unit 240 is shown in Fig. 5.

With reference to Fig. 5, the packet alignment unit 240 has a queue 241 provided corresponding to each unit switch at the first stage. The queue 241 is slotted on a packet basis. When a packet arrives, the packet alignment unit 240 identifies a unit switch at the first stage to which the packet in question is applied based

001100"06604560

5

10

15

20

25

Art.

A read pointer for reading a packet from the queue 241 is sequentially updated for each slot. Updating timing is assumed to be at packet reading or when no packet arrives for a time period longer than a predetermined time.

There, however, occurs a case where a packet assigned a sequence number (SN) at a unit switch at the first stage is abandoned before reaching a unit switch at the final stage due to buffer overflow of a unit switch on the way.

In this case, waiting for the packet in question to arrive will result in making a determination, when a following packet having the same sequence number (SN) as that assigned to the abandoned packet in question arrives after one circulation of the sequence numbers (SN), that the abandoned packets in question has arrived to conduct erroneous alignment of packets.

With reference to Fig. 7, shown is a state where with two packets assigned the sequence number "3", the packet first assigned the sequence number "3" is abandoned, whereby the packet later assigned the sequence number "3" is aligned together with a group of packets applied earlier in place of the abandoned packet.

For preventing such a situation, it is necessary to conduct processing of calculating an initial value of a read pointer based on a sequence number (SN) of a

00540990 033100

first received packet, as well as updating the read pointer under appropriate conditions. A method of determining an initial value of a read pointer and a method of updating the read pointer are shown in the flow charts of Figs. 8 and 9.

With reference to Fig. 8, operation of determining an initial value of a read pointer will be described. First, when a first packet arrives, determine whether a sequence number (SN) of the packet in question is an initial value (the number first assigned to a packet: assuming here that the initial value is "0") (Steps 801 and 802). When the sequence number (SN) is "0", that is, the initial value, set the value of the read pointer of a slot to "0" (Step 803) to finish processing (initial value of the read pointer is settled).

On the other hand, when the sequence number (SN) of the packet arriving first is not "0", set the value of the read pointer to the sequence number (SN) in question and set a timer (Step 804). Then, when none of new packets arrives before the timer in question overflows, finish processing with the value of the read pointer set at the Step 804 as an initial value (Step 805).

When a new packet arrives before the timer overflows (Steps 805 and 806), compare the sequence number (SN) of the packet in question and the value of

09540990-033100

5

10

15

20

25

When there exists no packet having a sequence number (SN) larger than the value of the read pointer, finish packet reading processing without updating the read pointer.

10

20

25

number of the outputs are four, three packets are sent out from the queue 241 at the uppermost stage and one packet is sent out from the queue 241 at the second stage.

5 In general, when there exist a number N of outputs, a number N of packets can be sent in one time slot. Then, each queue 241 tries to update the read pointer so as to output as many packets as possible. Accordingly, at each queue 241, read pointer is updated
10 a maximum of N times within one time slot.

As described in the foregoing, in a packet switch having unit switches connected in multi-stages, a unit switch at the first stage assigns a sequence number (SN) to an input packet and a unit switch at the final stage
15 outputs packets according to sequence numbers (SN) assigned to the received packets, which processing enables output of a packet which is to be output as soon as it arrives. It is accordingly unnecessary to cause a delay of a maximum time period required for a packet to
20 pass from a unit switch at the first stage to a unit switch at the final stage though a packet already arrives at the unit switch at the final stage as is done in a case where packets are sequenced using a time stamp.

Moreover, since packet output is executed
25 according to the order of sequence numbers (SN), packet alignment only needs determination whether a sequence number of an arriving packet is the same as that of a

001100 06604560

5

10

15

20

25

5

10

15

20

25

Next, description will be made of the prevention method by using a response message (ACK message). In an ACK message, a position of a read pointer is described. Here, a position of a read pointer represents a value subsequent to a sequence number (SN) of a packet whose

alignment processing is conducted last, that is, a value of a sequence number (SN) of a packet to be aligned next. The unit switch at the first stage updates sequence numbers (SN) up to modulo ($SN_{\text{currentRP}} + SN_{\text{length}}$) based on a value " $SN_{\text{currentRP}}$ " described in an ACK message without receiving an ACK message, assigns the updated sequence numbers to packets and transmits the packets.

The following packets will be kept waiting or abandoned. Such control enables a number " SN_{length} " of packets at the maximum to be accumulated at the unit switch at the second stage. Here, " SN_{length} " may vary for each combination between a unit switch at the input side (first stage) and a unit switch at the output side (final stage). At this time, " SN_{length} " can be considered to be a virtual queue length in the unit switch at the second stage as a whole.

The packet switch of the present embodiment is allowed to accommodate high-speed lines by aligning packets using sequence numbers (SN) as described above. Then, adopting the following manner as a method of assigning sequence numbers (SN) enables accommodation of a line whose speed is higher than that of an input port of a unit switch.

Assume that a port speed of a unit switch is 600 Mbps. Accommodate a line of 2.4 Gbps in this unit switch. In this case, as illustrated in Fig. 10, divide the line of 2.4 Gbps into four 600 Mbps lines on a packet basis.

5 conforming to the relevant rules, the order of separated packets can be also preserved. By restoring the packets in question according to the order of the sequence numbers (SN), the separated packets can be accurately restored.

10 In a manner as described above, a 2.4 Gbps line
can be accommodated using a unit switch equivalent to
600 Mbps. Similarly, OC-192 (10Gbps) can be accommodated
through division into 600 Mbps each. In a case where a
line of 2.4Gbps is accommodated using a unit switch
15 equivalent to 150 Mbps, the line in question is divided
into 16 lines of 150 Mbps on a packet basis. In a case
where the line of 2.4 Gbps is accommodated using a unit
switch equivalent to 1.2 Gbps, the line in question is
divided into two lines of 1.2 Gbps on a packet basis.

20 Thus, it is possible to accommodate a line whose speed
is higher than that of a port of a unit switch.

Fig. 11 is a block diagram showing a structure of a packet switch according to another embodiment of the present invention. Fig. 12 is a block diagram showing a structure of a unit switch in the packet switch of the present embodiment. With reference to Figs. 11 and 12, the present embodiment employs, as a unit switch, a $N^2 \times N^2$

packet switch 1120 constituted by connecting $2N \times 2N$ packet switches 1110 in two stages which have the same structure as that of the unit switch shown in Fig. 2. Then, a packet switch of $N^3 \times N^3$ is formed by further preparing a number N of the unit switches 1120 (#0 ~ #N-1) for extension.

In the present embodiment, in each packet switch 1110 constituting the unit switch 1120, a packet distribution unit 1111 and a packet alignment unit 1112 are prepared to conduct packet distribution and alignment. Since packet alignment operation using sequence numbers (SN) at the individual packet switch 1110 is the same as the packet alignment operation by the unit switch shown in Fig. 2 which has been described in the above first embodiment, no description will be made of the operation here.

Although the present invention has been described with respect to the preferred embodiments in the foregoing, the present invention is not always limited to the above-described embodiments. While in the above embodiments, the present invention is applied, for example, to a folded-type multi-stage switch, it is clearly understood that the present invention can be applied also to a multi-stage switch having the structure illustrated in Fig. 13.

As described in the foregoing, since according to the packet switch of the present invention, a sequence

00540990-033100

number is assigned to an input packet at a unit switch at a first stage and the packets are sequenced and output based on the sequence numbers assigned to the packets received at a unit switch at a final stage, it is unnecessary to delay a packet for a predetermined time period which is necessary in a case where packets are sequenced using a time stamp. This enables a time required for a packet to pass through a packet switch to be reduced, thereby improving quality of packet communication.

Further effect is enabling accommodation of a line whose speed is higher than that of an input port of a unit switch.

Moreover, securing a port for extension at a unit switch in advance facilitates expansion and extension of unit switches.

Furthermore, being structured as a folded-type multi-stage switch, the packet switch of the present invention realizes return of control information from a unit switch at a succeeding state to a unit switch at a preceding stage by the same control as that of a packet communication path, which enables control information returning processing to be executed with ease.

Although the invention has been illustrated and described with respect to exemplary embodiment thereof, it should be understood by those skilled in the art that the foregoing and various other changes, omissions and

001100 03100 06604560

additions may be made therein and thereto, without departing from the spirit and scope of the present invention. Therefore, the present invention should not be understood as limited to the specific embodiment set out above but to include all possible embodiments which can be embodied within a scope encompassed and equivalents thereof with respect to the feature set out in the appended claims.

007220" 06604560